



Identifying Influential and Susceptible Members of Social Networks

Sinan Aral and Dylan Walker

Science **337**, 337 (2012);
DOI: 10.1126/science.1215842

Identifying Influential and Susceptible Members of Social Networks

Sinan Aral* and Dylan Walker*

Identifying social influence in networks is critical to understanding how behaviors spread. We present a method that uses in vivo randomized experimentation to identify influence and susceptibility in networks while avoiding the biases inherent in traditional estimates of social contagion. Estimation in a representative sample of 1.3 million Facebook users showed that younger users are more susceptible to influence than older users, men are more influential than women, women influence men more than they influence other women, and married individuals are the least susceptible to influence in the decision to adopt the product offered. Analysis of influence and susceptibility together with network structure revealed that influential individuals are less susceptible to influence than noninfluential individuals and that they cluster in the network while susceptible individuals do not, which suggests that influential people with influential friends may be instrumental in the spread of this product in the network.

Peer effects are empirically elusive in the social sciences. Scholars in disciplines as diverse as economics, sociology, psychology, finance, and management are interested in whether children's peers influence their education; whether workers' colleagues influence their productivity; whether happiness, obesity, and smoking are "contagious"; and whether risky behaviors spread via peer influence. Answers to these questions are critical to policy because the success of intervention strategies in these domains depends on the robustness of estimates of the degree to which contagion is at work during a social epidemic (*1*, *2*). Robust estimation of peer effects is also critical to understanding whether new social media technologies magnify peer influence in product demand, voter turnout, and political mobilization or protest.

The recent availability of population-scale networked data sets generated by e-mail, instant messaging, mobile phone communications, and online social networks enables novel investigations of the diffusion of information and influence in networks (*3*–*9*). Unfortunately, identifying influence in these networks is difficult because estimation is confounded by homophily [the tendency for individuals to choose friends with similar tastes and preferences (*10*, *11*), and thus for preferences to be correlated among friends], confounding effects (the tendency for connected individuals to be exposed to the same external stimuli), and simultaneity (the tendency for connected individuals to co-influence each other and to behave similarly at approximately the same time), among other factors (*1*, *2*, *10*, *12*–*17*). Although some new methods separate peer influence from homophily and confounding factors in observational data (*11*), controlling for unobservable factors such as latent homophily (correlation among unobserved drivers of preferences among friends) remains difficult without exogenous variation in adoption probabilities across individuals (*18*). Fortunately, randomized experiments provide a more robust means of identifying causal peer effects in networks (*19*–*22*).

One particularly controversial argument in the peer effects literature is the "influentials" hypothesis—the idea that influential individuals catalyze the diffusion of opinions, behaviors, innovations, and products in society (*23*, *24*). Despite the popular appeal of this argument, a variety of theoretical models suggest that susceptibility, not influence, is the key trait that drives social contagions (*25*–*29*). Little empirical evidence exists to adjudicate these claims. Understanding whether influence, susceptibility to influence, or a combination of the two drives social contagions, and accurately identifying influential and susceptible individuals in networks, could enable new behavioral interventions to affect obesity, smoking, exercise, fraud, and the adoption of new products and services.

We conducted a randomized experiment to measure influence and susceptibility to influence in the product adoption decisions of a representative sample of 1.3 million Facebook users. The experiment involved the random manipulation of influence-mediating messages sent from a commercial Facebook application that lets users share information and opinions about movies, actors, directors, and the film industry. As users adopted and used the product, automated notifications of their activities were delivered to randomly selected peers in their local social networks. For example, when a user rated a movie on the application, a randomly selected subset of the user's Facebook friends were sent a message notifying them of the rating with a link to the canvas page describing the application and instructions on how to adopt it. Because message recipients were randomly selected, treated and untreated peers of the application user differed only by the number of randomized messages they received. Estimates of influence and susceptibility were obtained by modeling time to peer adoption as a function of the peer's treatment status—whether influence-mediating messages had been received, and if so, how many. An influence-mediating message generally refers to any communication between peers that could conduct influence (*19*, *30*), such as wearing a logo advertising a brand or recommending a product to a friend.

The experiment was conducted over 44 days during which 7730 product adopters sent 41,686 automated notifications to randomly chosen targets among their 1.3 million friends. This resulted in 976 unique peer adoptions, or a 13% increase in demand for the product relative to the number of initial adopters (see tables S1 to S4 and figs. S1 to S4).

Our method avoids several known sources of bias in influence identification by randomly manipulating who receives influence-mediating messages. First, we avoid selection bias by randomizing whether and to whom influence-mediating messages are sent (table S5). In uncontrolled environments, users may choose to send messages to peers who are more likely to like the product or to listen to their advice, which confounds estimates of susceptibility to influence by oversampling recipients who are more likely to respond positively. Second, our method eliminates bias created by homophily or assortativity in networks by randomizing the receipt of influence-mediating messages. Even latent homophily is controlled because similarity in unobserved attributes is equally represented across treatment groups. Third, the method controls for unobserved confounding factors, because randomly chosen peers are equally likely to be exposed to external stimuli that affect adoption (such as advertising campaigns or promotions). Fourth, the automatically generated messages include identical information, eliminating heterogeneity in message content and valence, which are known to affect responses to social influence (*31*). Differences in adoption between treatment groups can then be attributed solely to the number of influence-mediating messages they received.

Our statistical approach used hazard modeling, which is the standard technique for estimating social contagion in economics, marketing, and sociology [e.g., (*32*)]. However, we extended existing techniques to distinguish two types of peer adoption: (i) spontaneous adoption, which occurs in the absence of influence, and (ii) influence-driven adoption, which occurs in response to persuasive messages. This extension is important because even in the absence of influence, adoption outcomes cluster among peers as a consequence of homophily, assortativity, simultaneity, and correlated effects (*11*, *12*). We estimate the average treatment effects of notifications by aggregating many individual experiments in which messages were randomized within the local networks of the original adopting users (tables S6 and S7).

To estimate the moderating effects of an individual *i*'s attributes on the influence exerted

Stern School of Business, New York University, New York, NY 10012, USA.

*To whom correspondence should be addressed. E-mail: sinan@stern.nyu.edu (S.A.); dwalker@stern.nyu.edu (D.W.)

by i on peer j (and to distinguish them from the moderating effects of j's attributes on j's susceptibility to influence), we use a continuous-time single-failure proportional hazards model. Survival models provide information about how quickly peers respond (rather than simply whether they respond) and also correct for censoring of peer responses that may occur beyond the experiment's observation window. We specify the following model:

$$\lambda_j(t, X_i, X_j, N_j) = \lambda_0(t)\ \exp[N_j(t)\beta_N + X_i\beta^i_{spont} + X_j\beta^j_{spont} + N_j(t)X_i\beta_{infl} + N_j(t)X_j\beta_{susc}] \quad (1)$$

where λ_j is the hazard of peer j of application user i adopting the application (each peer j is associated with one and only one application user i), $\lambda_0(t)$ represents the baseline hazard, X_i represents a set of individual attributes of an application user i, X_j represents a set of individual attributes of peer j, $N_j(t)$ represents the number of notifications received by peer j of application user i as a function of time, $N_j(t)$ reflects the extent to which j has been exposed to influence-mediating messages from user i, β_N estimates the effect of receiving a notification on the likelihood of peer adoption (holding sender and potential recipient attributes constant), β^i_{spont} estimates the propensity for peers of user i with attributes X_i to spontaneously adopt in the absence of influence ($N_j = 0$), β^j_{spont} estimates the propensity for peer j with attributes X_j to spontaneously adopt in the absence of influence ($N_j = 0$), β_{infl} estimates the impact of user i's attributes on i's ability to influence peer j to adopt the application above and beyond j's propensity to adopt spontaneously, and β_{susc} estimates the impact of j's attributes on j's likelihood to adopt as a result of influence above and beyond j's propensity to adopt spontaneously (for alternative specifications, robustness, and goodness of fit, see table S8 and figs. S5 to S12).

Models of dyadic (two-party) relationships between influencers and potential susceptibles test whether influence depends on characteristics of the relationship between a given pair—for example, whether women are more influential over men than men are over women. To estimate the effect of dyadic relationships, we use the following continuous-time single-failure proportional hazards model:

$$\lambda_j(t, X_i, X_j, N_j) = \lambda_0(t)\ \exp[N_j(t)\beta_N + S(X_i, X_j)\beta^{i-j}_{spont} + N_j(t)S(X_i, X_j)\beta^{i\rightarrow j}_{infl}] \quad (2)$$

where X_i represents a set of the individual attributes of the sender, X_j represents a set of the individual attributes of peer j (the potential recipient), and $S(X_i, X_j)$ represents a set of dyadic covariates that characterize the joint attributes of the sender-recipient pair. Dyadic covariates estimate, for example, whether influence is stronger when the sender and recipient are the same or different genders. β_{spont} estimates the effect of a dyadic relationship between application user i and peer j on the tendency for j to adopt spontaneously. For example, when the dyadic relationship variable is an indicator of similarity (such as "same age"), β_{spont} captures the extent to which similarity on that dimension predicts the likelihood to spontaneously adopt, and represents the propensity to adopt as a result of preference similarity and other explanations for correlations in adoption likelihoods between peers that are not a result of influence. β_{infl} estimates the effect of the dyadic attribute (e.g., "same age") on the degree to which i influences j to adopt, above and beyond j's likelihood to spontaneously adopt.

On average, susceptibility decreases with age (Fig. 1). People over the age of 31 are the least susceptible to influence; relative to people who do not declare their age, they have an 18% lower hazard of adopting the application upon receiving a notification ($P < 0.05$; the statistical significance of all estimates is derived from χ^2 tests). However, people in this same age quartile (>31) are significantly more influential than people in the lowest age quartile (<18). Relative to people younger than 18, people over 31 have a 51% greater instantaneous likelihood of influencing their peers to adopt with an influence-mediating message ($P < 0.05$).

Men are 49% more influential than women ($P < 0.05$), but women are 12% less susceptible to influence than men ($P < 0.05$). Single and married individuals are the most influential. Single individuals are significantly more influential than those who are in a relationship (113% more influential, $P < 0.05$) and those who report their relationship status as "It's complicated" (128% more influential, $P < 0.05$). Married individuals are 140% more influential than those in a relationship ($P < 0.01$) and 158% more influential than those who report that "It's complicated" ($P < 0.01$). Susceptibility increases with increasing relationship commitment until the point of marriage. People who are engaged to be married are 53% more susceptible to influence than single people ($P < 0.05$), whereas married individuals are the least susceptible to influence ($P = 0.93$, n.s.). The engaged and those who report that "It's complicated" are the most susceptible to influence. Those who report that "It's complicated" are 111% more susceptible to influence than baseline users who do not report their relationship status on Facebook ($P < 0.05$), and those who are engaged are 117% more susceptible than baseline users ($P < 0.001$).

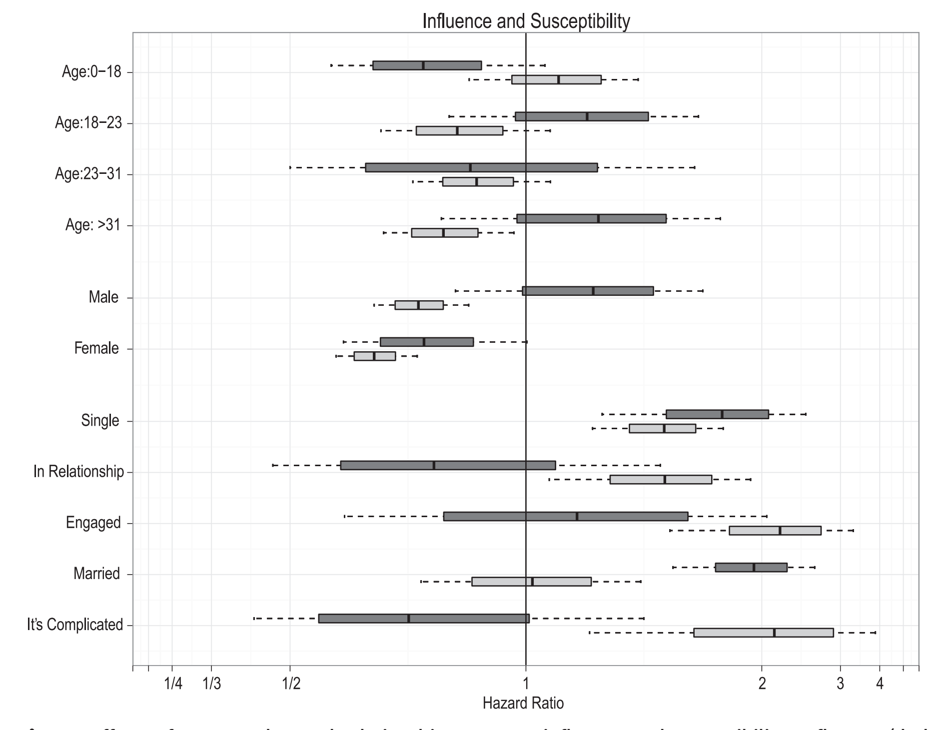


Fig. 1. Effects of age, gender, and relationship status on influence and susceptibility. Influence (dark gray) and susceptibility to influence (light gray) are shown with SEs (boxes) and 95% confidence intervals (whiskers). The figure displays hazard ratios (HRs) representing the percent increase (HR > 1) or decrease (HR < 1) in adoption hazards associated with each attribute. Age is binned by quartiles. Each attribute is shown as a pair of estimates, one reflecting influence (dark gray) and the other susceptibility (light gray). Personal relationship status reflects the status of an individual's current romantic relationship and is specified on Facebook as Single, In a Relationship, Engaged, Married, or It's Complicated. Estimates are shown relative to the baseline case for each attribute, which is the average for individuals who do not display that attribute in their online profile.

People exert the most influence on peers of the same age [97% more influence than baseline ($P < 0.01$)] (Fig. 2). They also seem to exert more influence on younger peers than on older peers, although this difference is not significant. In nondyadic susceptibility models, we found that women were less susceptible to influence than men (Fig. 1). Dyadic models (Fig. 2) further revealed that women exert 46% more influence over men than over other women ($P = 0.01$). Finally, individuals in equally (and more) committed relationships relative to their peers (e.g., those who are married versus those who are engaged, in a relationship, or single) are significantly more influential [equally committed, 70% more influential than baseline ($P < 0.05$); more committed, 101% more influential than baseline ($P < 0.05$)], although future work will be needed to determine whether there is something "different" about people who do not provide some information (e.g., age) (table S1).

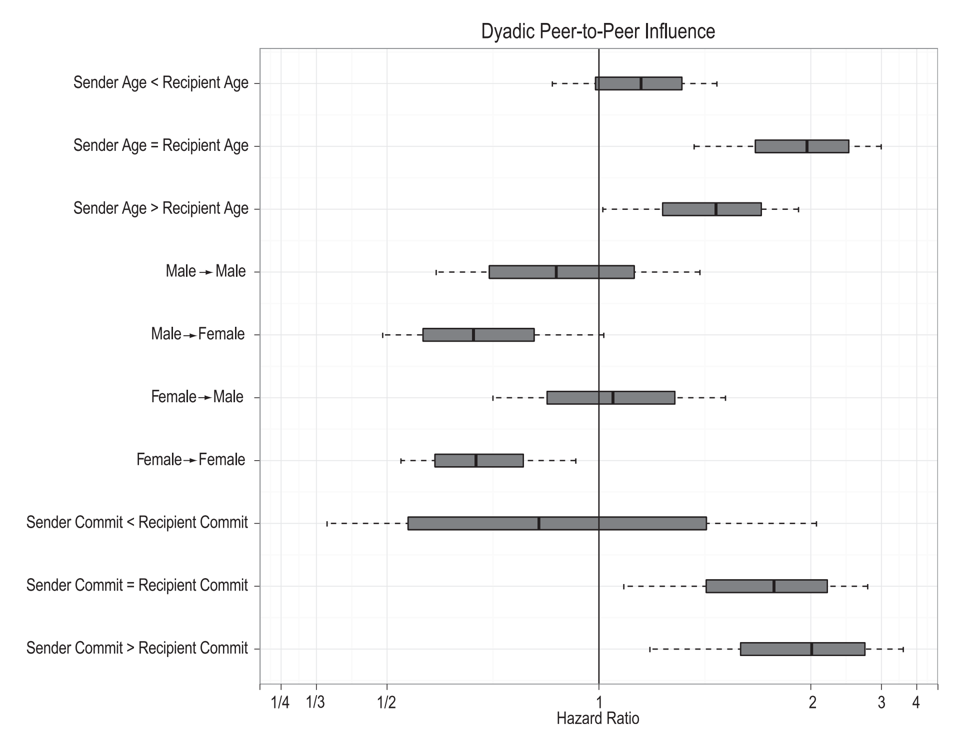


Fig. 2. Dyadic influence models involving age, gender, and relationship status. The results include the relative age, gender similarity, and commitment level of the relationship status of senders and recipients, with SEs (boxes) and 95% confidence intervals (whiskers). The figure displays hazard ratios representing the percent increase (HR > 1) or decrease (HR < 1) in adoption hazards associated with each attribute. The baseline case represents dyads in which the attribute being examined is unreported in the Facebook profile of one or both peers.

Comparing spontaneous adoption hazards to influenced adoption hazards reveals the potential roles that different individuals play in the diffusion of a behavior (Fig. 3). For example, in the case of the movie product we studied, both single and married individuals adopt spontaneously more often than baseline users [single, 31% more often ($P < 0.05$); married, 36% more often ($P = 0.06$)], are more influential than baseline users [single, 71% more influential ($P < 0.01$); married, 94% more influential ($P < 0.001$); Fig. 1], and have peers who are no more likely to adopt spontaneously than baseline users ($P = 0.39$ and 0.08; n.s.). This suggests that influence exerted by single and married individuals positively contributes to this product's diffusion without any need to target their peers. On the other hand, women are poor candidates for targeted advertising because they are likely to adopt spontaneously and are 22% less influential on their peers than baseline users ($P < 0.05$). Those who claim that their relationship status is complicated are easily influenced by their peers to adopt [35% more susceptible than baseline ($P < 0.05$)] but are not influential enough to spread the product further ($P = 0.49$; n.s.). These results have implications for policies designed to promote or inhibit diffusion, and they illustrate the general utility of our method for informing intervention strategies, targeted advertising, and policy-making.

Figure 4 shows the joint distributions of influence and susceptibility in a network, reveal-

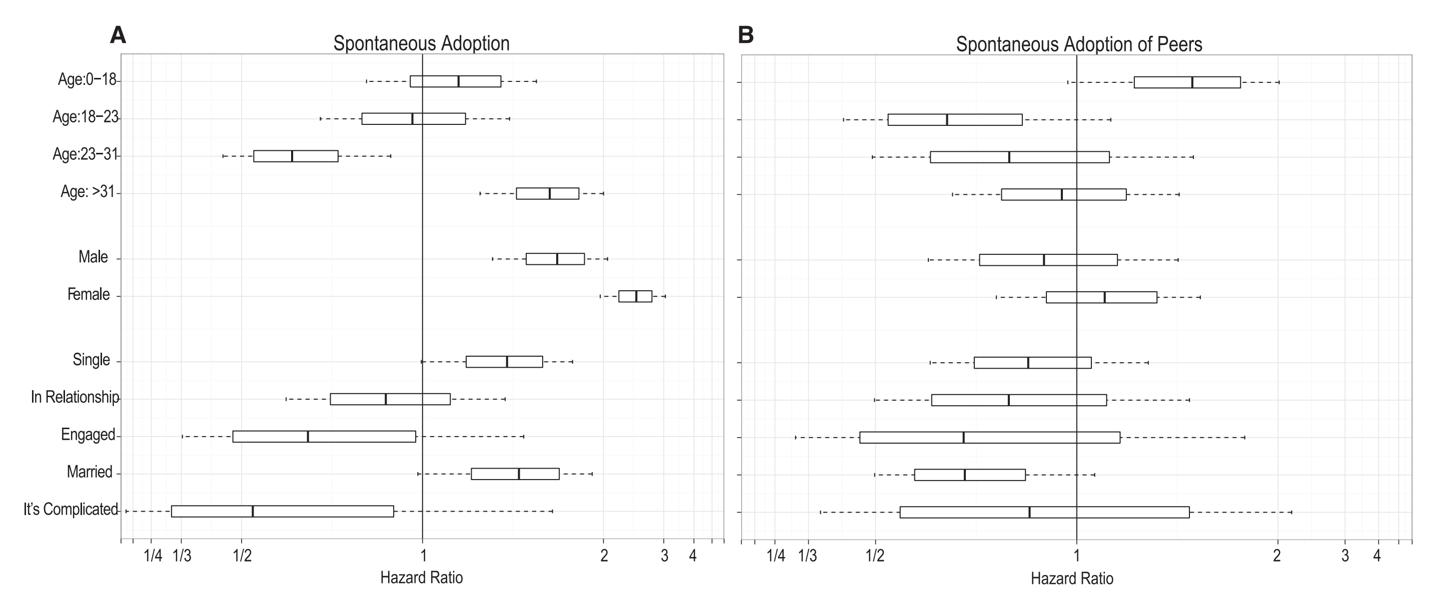


Fig. 3. (**A**) Hazard ratios for individuals to adopt spontaneously as a function of their attributes, with SEs (boxes) and 95% confidence intervals (whiskers). (**B**) Hazard ratios for individuals to have peers who adopt spontaneously as function of their attributes. The figure displays hazard ratios representing the percent increase (HR > 1) or decrease (HR < 1) in adoption hazards associated with each attribute.

ing the correlation of influence and susceptibility across all individuals and the assortativity of influence and susceptibility across all individuals and their peers in the network. We calculated individual influence and susceptibility scores as the product of the estimated hazard ratios of individuals' attributes for a broader sample of 12 million users with 85 million relationships. The analysis combines the estimated impact of each demographic attribute on influence and susceptibility to calculate individuals' overall influence and susceptibility scores. For example, a 35-year-old single female has an influence score equal to $\exp(\beta_{\text{infl, >31}} + \beta_{\text{infl, single}} + \beta_{\text{infl, female}})$. The following inferences can be drawn from our results:

1) Highly influential individuals tend not to be susceptible, highly susceptible individuals tend not to be influential, and almost no one is both highly influential and highly susceptible to influence (Fig. 4, panel I). This implies that influential individuals are less likely to adopt the product as a consequence of natural influence processes (i.e., in the absence of targeting); hence, targeting influentials with low propensities to spontaneously adopt would be a potentially viable promotion strategy.

2) The "influentials" and "susceptibles" hypotheses are orthogonal claims. Both influential individuals and noninfluential individuals have approximately the same distribution of susceptibility to influence among their peers; hence, being influential is not simply a consequence of having susceptible peers (Fig. 4, panel II). Both influence and susceptibility play a role in the peer-to-peer diffusion of the product. Combining studies of influence with studies of susceptibility will therefore likely improve our understanding of the diffusion of behavioral contagions.

3) There are more people with high influence scores than high susceptibility scores (Fig. 4, panel I), which suggests that, in our context, targeting should focus on the attributes of current adopters (e.g., giving individuals incentives to influence their peers) rather than attributes of their peers (e.g., giving individuals with susceptible peers incentives to adopt).

4) Influentials cluster in the network. As shown in Fig. 4, panel III, influential individuals connected to other influential peers are approximately twice as influential as baseline users. In contrast, we find a tendency for less susceptible users to cluster together and no clusters of highly susceptible users (Fig. 4, panel IV). The clustering of influentials suggests the existence of a multiplier effect of infecting a highly influential individual. However, such individuals also tend to have peers with only average susceptibility, making predictions about which effect would dominate difficult without more evidence. Additional empirical and simulation studies should therefore examine the effects of the assortativity of influence and susceptibility on the diffusion of behaviors, products, and diseases.

Analyzing the heat maps in Fig. 4 is not sufficient to identify optimal intervention targets, because more information is needed about the network structure around candidate targets in each region. For example, an individual with high influence and high peer susceptibility in the upper right quadrant of in Fig. 4, panel II, may seem like a good target, but may be of low degree or may be isolated. The network diagrams to the left of the heat maps show the assortativity of influence and susceptibility in ego networks from different regions combined with information on their network structures, such as network degree and the distribution of influence and susceptibility across peers in the network. Analyzing networks in different regions of the

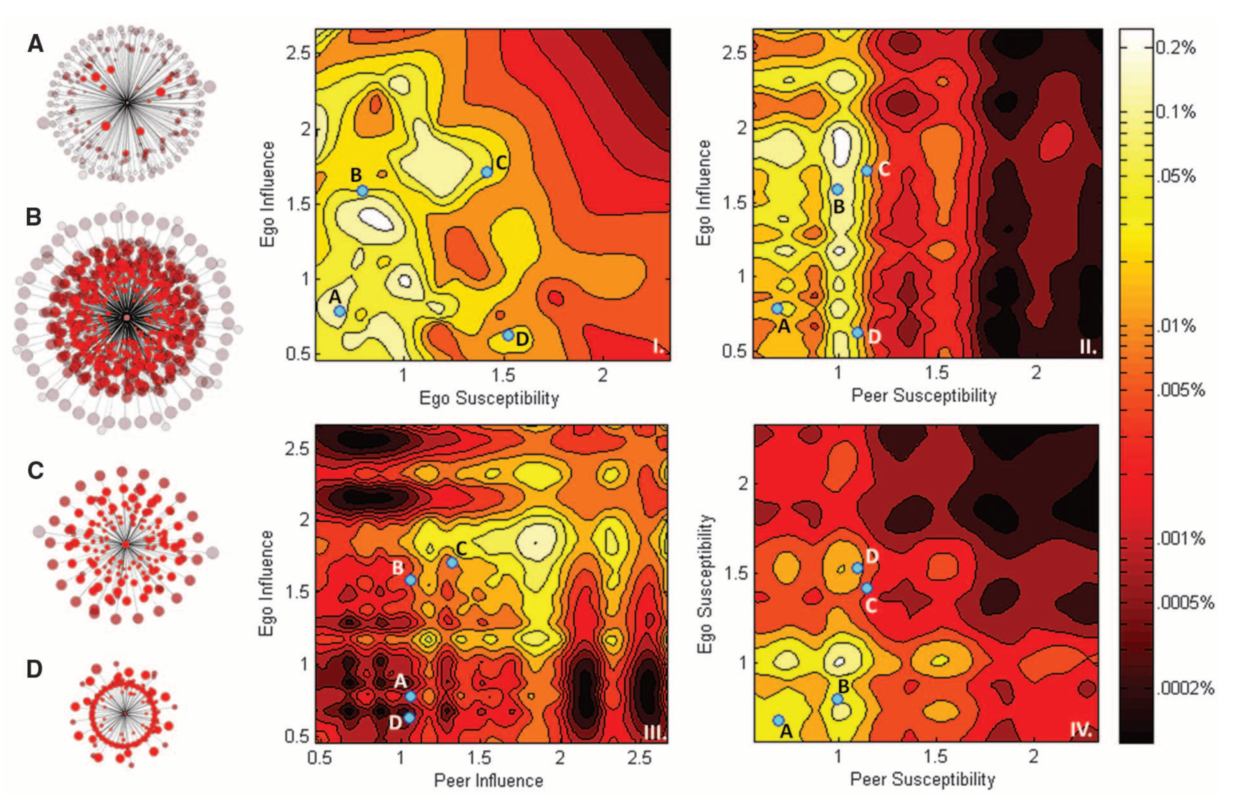


Fig. 4. Scores for 12 million Facebook users (collected from users who installed one of several other Facebook applications developed by the company) with 85 million relationships are calculated by means of hazard rate estimates relative to the baseline hazard in the influence and susceptibility model described in the text. The resulting heat maps are shown at the right. Panel I displays the percentage of people (ego) with predicted influence (*y* axis) and predicted susceptibility (*x* axis). Panels II to IV display the percentage of ego-peer relationships: panel II, ego influence (*y* axis) and peer susceptibility (*x* axis); panel III, ego influence (*y* axis) and peer influence (*x* axis); and panel IV, ego susceptibility (*y* axis) and peer susceptibility (*x* axis). The heat maps do not provide information on network structure, which can be important for informing targeting decisions. The diagrams to the left of the heat maps show the assortativity of influence and susceptibility in ego networks drawn from the regions of the heat maps labeled A, B, C, and D. Nodes in the networks are sized in proportion to their predicted influence (larger nodes are more influential) and are shaded and placed relative to their predicted susceptibility (redder nodes and nodes closer to ego are more susceptible; grayer nodes and nodes farther from ego are less susceptible).

heat maps, such as those displayed in Fig. 4, can suggest optimal targets. For example, node C is not only highly influential, highly susceptible, and has peers who are themselves influential and susceptible, but is also of above average degree in its region and has many peers who are susceptible rather than one highly susceptible peer driving the average susceptibility in its network. These characteristics in combination make C a good target.

Our method uses randomized experiments to identify influential and susceptible individuals in large social networks; however, the work does have limitations. Although we avoid bias by randomizing message recipient selection and holding message content constant, recipient selection and message content may be important aspects of influence and should therefore be estimated in future experiments. Furthermore, it is still not clear whether influence and susceptibility are generalized characteristics of individuals or instead depend on which product, behavior, or idea is diffusing. Although our estimates should generalize to the diffusion of similar products, they are not conclusions about who is more or less influential in general. Our experimental methods for influence identification, however, are generalizable and can be used to measure influence and susceptibility in the diffusion of other products and behaviors in a variety of settings.

Previous research has taken an individualistic view of influence—that someone's importance to the diffusion of a behavior depends only on his or her individual attributes or personal network characteristics. In contrast, our results show that the joint distributions of influence, susceptibility, and the likelihood of spontaneous adoption in the local network around individuals together determine their importance to the propagation of behaviors. Future research should therefore examine how the codistribution of influence, susceptibility, and dyadic induction in networks affects the diffusion of behaviors, the development of social contagions, and the effects of policies intended to promote or contain behavior change. More generally, our results show the potential of methods based on large-scale in vivo randomized experiments to robustly estimate peer effects and identify influential and susceptible members of social networks.

References and Notes

1. S. Aral, *Mark. Sci.* **30**, 217 (2011).
2. S. Aral, D. Walker, *IEEE Intell. Syst.* **26**, 91 (2011).
3. G. Kossinets, D. J. Watts, *Science* **311**, 88 (2006).
4. D. Lazer *et al.*, *Science* **323**, 721 (2009).
5. N. Eagle, M. Macy, R. Claxton, *Science* **328**, 1029 (2010).
6. S. A. Golder, M. W. Macy, *Science* **333**, 1878 (2011).
7. S. Aral, M. W. Van Alstyne, *Am. J. Sociol.* **117**, 90 (2011).
8. E. Sun, I. Rosenn, C. Marlow, T. Lento, in *Proceedings of the Third International Conference on Weblogs and Social Media* (AAAI Press, Menlo Park, CA, 2009).
9. J. Leskovec, L. A. Adamic, B. A. Huberman, *ACM Trans. Web* 10.1145/1232722.1232727 (2007).
10. M. McPherson, L. Smith-Lovin, J. M. Cook, *Annu. Rev. Sociol.* **27**, 415 (2001).
11. S. Aral, L. Muchnik, A. Sundararajan, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21544 (2009).
12. C. F. Manski, *Soc. Methodol.* **23**, 1 (1993).
13. S. Currarini, M. O. Jackson, P. Pin, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4857 (2010).
14. H. Noel, B. Nyhan, *Soc. Networks* **33**, 211 (2011).
15. W. Hartmann *et al.*, *Mark. Lett.* **19**, 287 (2008).
16. N. A. Christakis, J. H. Fowler, *N. Engl. J. Med.* **357**, 370 (2007).
17. R. Lyons, *Stat. Polit. Policy* 10.2202/2151-7509.1024 (2011).
18. C. R. Shalizi, A. C. Thomas, *Sociol. Methods Res.* **40**, 211 (2011).
19. S. Aral, D. Walker, *Manage. Sci.* **57**, 1623 (2011).
20. D. Centola, *Science* **329**, 1194 (2010).
21. S. Leider, M. M. Möbius, T. Rosenblat, Q.-A. Do, *Q. J. Econ.* **124**, 1815 (2009).
22. E. Bakshy, I. Rosenn, C. Marlow, L. Adamic, in *WWW '12 Proceedings of the 21st International Conference on World Wide Web* (ACM, New York, 2012), pp. 519–528.
23. E. Katz, *Public Opin. Q.* **21**, 61 (1957).
24. T. W. Valente, *Network Models of the Diffusion of Innovations* (Hampton, Cresskill, NJ, 1995).
25. D. Kempe, J. Kleinberg, É. Tardos, in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York, 2003), pp. 137–146.
26. M. Granovetter, *Am. J. Sociol.* **83**, 1420 (1978).
27. D. J. Watts, P. S. Dodds, *J. Consum. Res.* **34**, 441 (2007).
28. P. S. Dodds, D. J. Watts, *Phys. Rev. Lett.* **92**, 218701 (2004).
29. D. Centola, M. Macy, *Am. J. Sociol.* **113**, 702 (2007).
30. D. Godes *et al.*, *Mark. Lett.* **16**, 415 (2005).
31. C. Heath, C. Bell, E. Sternberg, *J. Pers. Soc. Psychol.* **81**, 1028 (2001).
32. R. Iyengar, C. Van den Bulte, T. W. Valente, *Marketing Sci.* **30**, 195 (2011).

Acknowledgments: We thank S. Aral, H. Frydman, C. Hurvich, P. Perry, J. Simonoff, and M. Sternberg for invaluable discussions. Supported by a Microsoft research faculty fellowship (S.A.) and by NSF Career Award 0953832 (S.A.). The research was approved by the NYU institutional review board. There are legal obstacles to making the data available, but code is available upon request. The requests for data and randomization of message targets we used are standard ways in which applications request and use user data on Facebook. They are covered by the Facebook privacy policy and terms of service. Opt-in permissions were granted by the user to the application developer on a per-application basis when the user installed the application, via Facebook application authentication dialogs. In the dialogs we asked for all the categories of data we used in the study, and all of these requests were in line with the Facebook terms of service. Users saw these requests and opted in to them before installing the app.

Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1215842/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S9

26 October 2011; accepted 30 May 2012
Published online 21 June 2012;
10.1126/science.1215842

Sex-Specific Adaptation Drives Early Sex Chromosome Evolution in *Drosophila*

Qi Zhou and Doris Bachtrog*

Most species' sex chromosomes are derived from ancient autosomes and show few signatures of their origins. We studied the sex chromosomes of *Drosophila miranda*, where a neo-Y chromosome originated only approximately 1 million years ago. Whole-genome and transcriptome analysis reveals massive degeneration of the neo-Y, that male-beneficial genes on the neo-Y are more likely to undergo accelerated protein evolution, and that neo-Y genes evolve biased expression toward male-specific tissues—the shrinking gene content of the neo-Y becomes masculinized. In contrast, although older X chromosomes show a paucity of genes expressed in male tissues, neo-X genes highly expressed in male-specific tissues undergo increased rates of protein evolution if haploid in males. Thus, the response to sex-specific selection can shift at different stages of X differentiation, resulting in masculinization or demasculinization of the X-chromosomal gene content.

X and Y chromosomes follow distinctive evolutionary trajectories after recombination becomes suppressed between ancestral homologous autosomes with a sex-determining function (*1*). The lack of recombination greatly impairs natural selection on the Y, which loses most of its original genes and often accumulates repetitive DNA (*2*). However, Y chromosomes are not complete evolutionary dead ends; instead, their male-limited transmission favors the gain of male-related genes ("masculinization"). Low gene density yet enrichment of male-specific genes is shared among many independently evolved ancient Ys (*3*, *4*), but few traces of their evolutionary origins remain, making processes involved in Y degeneration little understood. Conversely, the X still recombines in females, and selection can effectively purge deleterious alleles and incorporate beneficial mutations (*2*). Unlike autosomes, the X is transmitted more often through females than males, favoring an underrepresentation of male-beneficial genes on the X ("demasculinization") (*5*, *6*). Further, almost all X-linked genes are haploid in males (hemizygous) and can fix re-

Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: dbachtrog@berkeley.edu